2025 ASEE Northeast Section Conference, March 22, 2025, University of Bridgeport, Bridgeport, CT, USA.

A Novel Platform for Evaluating Nurses' Trust in AI-based Biomedical Tools

Boluwatife E. Faremi[1], Javier O. Pinzon-Arenas[1],
Amir Mohammad Karimi Forood[1], Josef Kundrát[1],
Hugo F. Posada-Quintero[1], Ann Marie Hoyt-Brennan[2], Wendy A. Henderson[2]
[1]Department of Biomedical Engineering, University of Connecticut, Storrs, CT
[2]Penn Nursing, University of Pennsylvania, Philadelphia, PA

***Abstract*—AI-based biomedical engineering tools are increasingly being used to address critical challenges in various aspects of healthcare, from diagnostics to therapeutics. While AI offers immense potential to improve clinical decision making, the trust of intended users (e.g., nurses and physicians) in these systems remains largely unexplored. This study addresses the gap in medical education by investigating how nursing students interact with and trust AI recommendations in realistic healthcare scenarios. In a multidisciplinary collaboration of experts in biomedical engineering, nursing, psychology, and simulation, we present a novel virtual platform for simulating healthcare tools to assess students' trust in AI recommendations using custom-designed scenarios illustrated with video vignettes. Using different AI systems with varying levels of performance and combinations of correct and incorrect suggestions, the platform is designed to provide an in-depth exploration of these trust dynamics in realistic healthcare settings. The platform allows the collection of participants' trust levels, cognitive loads, reaction times, and physiological reactions throughout the experiment using validated tools. Physiological measures, particularly electrodermal activity, aim to capture the effect of trust in the emotions of the participants during the study. The platform enables researchers to study how different AI performance and scenario complexity affect trust, decision making, and cognitive load for users, and can help inform the development of future targeted educational interventions aimed at optimizing the integration of AI into healthcare education and practice.**

Keywords—trust, performance, electrodermal activity, AI, healthcare, education, engineering

I. INTRODUCTION

AI is revolutionizing healthcare and medical education by enabling automated assessments, personalized learning, real-time content updates, clinical simulations, and adaptation of educational materials to reflect current research and practice [1]. This technological shift comes at a time when the traditional method of medical education faces significant challenges, such as limited hands-on experience, inconsistent mentoring, and the burden on students to memorize and replicate complex real-world scenarios taught in resource-constrained settings [2]. Trust in AI has been identified as essential to the successful adoption of AI in health professional education [3]. Trust in technology has been investigated from a psycho-physiological perspective across diverse domains from driving simulation to virtual reality and collaborative robotics [4], [5], [6]. Yet, a significant gap exists in the nursing field, as no study has addressed nurses' trust in biomedical AI tools. We present a platform for measuring and monitoring trainee nurses' trust in AI healthcare technology systems (AIHTs).

Existing approach in assessing trust in AI tools for learning and guiding trainees in patient outcomes has been largely dependent on expert supervisors, educators, or clinicians, who can swiftly determine the relevance and applicability of AI-generated outputs—an ability that novice learners may lack [1]. Regardless of expert validation, it has been declared that trainees may either underuse AI tools due to distrust or become overly reliant on them, potentially leading to professional deskilling [7]. Therefore, to effectively integrate AI into medical training, there's a critical need for AIHTs that can measure and monitor trainee trust in AI tools within healthcare education. This is essential because understanding trainee trust levels will allow educators to design tailored learning pathways, identify strengths, and weaknesses in knowledge and reinforce critical concepts before trainees interact with real patients [8].

Defining trust is complex due to variations in interpretation across different fields [9]. Nonetheless, in this study, trust is defined as the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party [10]. The complexity of human-AI interaction necessitates examining trust across three dimensions: dispositional, situational, and learned trust [5]. Dispositional trust is a stable tendency to trust AI, shaped by personality traits and past experiences. Situational trust, by contrast, is dynamic and context-dependent, influenced by system complexity, task difficulty, and perceived risk.

Learned trust develops through direct interactions with AI, updating over time based on experience. In healthcare, which is the focus of this study, trust measurement primarily relies on situational trust, as users assess AI based on real-time performance and perceived reliability. Nevertheless, across all types of trust, cognition plays a crucial role in shaping trust through beliefs, prior experiences, and continuous reassessment based on new information [11], [12], [13].

Several attempts have been made in literature to assess trust from a pyscho-physiological perspective since it carries a cognitive component [4]. Trusts manifestation in the human body is the result of a complex series of physiological events. One related connection can be seen by changes in a person's skin conductance because of changes in stress level associated with trust. Skin conductance changes captured by electrodermal activity (EDA) are not under conscious control but rather is altered by the sympathetic innervations of sweat glands, which cause increase on sweating production. It is of importance to mention that EDA is also called galvanic skin response. Continuous monitoring of EDA has been widely explored in many settings including monitoring pain, stress, fatigue, and trust [14], [15], [16]. Several recent studies have investigated the link between trust and EDA, showing that lower trust correlates with higher EDA. An early study showed that the EDA of users of a text chat environment was strongly affected by trust and cognitive load [17]. Recent studies have further validated the suitability of EDA to assess trust [18], [19], [20], [21].

Various approaches to measuring trust in AI and automated systems have been explored in the literature. While our study focuses on nurses interacting with AI healthcare tools, these methods build upon prior research in related domains. Trust in automative sensors was evaluated using EEG and EDA during simulated driving tasks [5]. Similarly, EDA and EEG were combined to assess trust and cognitive load in virtual reality environments, demonstrating the versatility of these physiological measures across different interaction modalities [4]. In the domain of robotics, EDA was specifically investigated as a primary physiological marker for indicating trust in human-robot collaboration [6]. Most relevant to our healthcare focus, the impact of enhanced factual explanations on trust in AI systems was examined by measuring both blood volume pressure and EDA [22]. Our AIHT platform extends these methodologies into the critical healthcare education context, where accurate trust calibration has direct implications for patient care, while preserving the proven effectiveness of EDA as a trust indicator across various human-AI interaction scenarios.

This paper aims to design a platform that addresses the identified gap in medical education by measuring and monitoring trainee nurses' trust in AIHT, by integrating a set of simulated scenarios and simultaneous collection of EDA signals. The AIHT platform aims to assess the relationship between trust and nurse-AIHT performance, reducing reliance on expert judgment alone for evaluating trust in AIHT technologies. Its potential benefit lies in fostering appropriate trust among trainees, enabling them to make critical decisions effectively in fast-paced real-life scenarios. The platform and preliminary validation of the platform are presented in the following section.

II. MATERIALS AND METHODS

In synthesis, The AIHT platform is an interactive system designed to assess nurses' decision-making and trust in AI-assisted healthcare scenarios. Participants watch realistic, custom-designed video scenarios where they must decide on a course of action—using an AED, administering NARCAN, or doing nothing—based on an AI system's recommendation. The platform measure's reaction time provides immediate feedback on decision accuracy and evaluates trust using the Human-Computer Trust Scale (HCTS) and workload demands via the NASA-TLX questionnaire. To examine the impact of AI performance on trust, the platform allows the participants to interact with both high-performance (HPAI) and low-performance (LPAI) AI systems across 40 scenarios, with randomized exposure order. The AIHT platform is illustrated in Fig. 1. A more detailed description of the AIHT platform is provided below.

AIHT Platform Design: The platform presents the participant a series of realistic scenarios illustrated with videos. The videos depict a realistic scenario illustrated with vignettes in which a nurse ultimately had to make a critical decision with the assistance of an AI system. A pool of 50 scenarios were designed. For this implementation, the participant has three options: automated external defibrillator (AED), administer NARCAN, or do nothing. These videos were custom-designed by members of the research team from the UConn School of Nursing to mimic nursing-related events that commonly incorporated AI technology. For example, in a given scenario a patient has a deteriorating condition and it's important to decide if cardiopulmonary resuscitation is needed. Given the specifics of the situation, the AI system recommends using the AED. Then, the participant must choose to follow the AIHT suggestion or choose another option. The participant receives a response notifying them of the "correctness" of their answer (the patient survived or not).

The AIHT platform also assesses human-computer trust using the HCTS [23], which measures perceptions of benevolence, competence, and perceived risk in human-technology interactions. Additionally, the AIHT platform includes the NASA-TLX questionnaire after each testing group to evaluate workload demands while performing the task [24]. Given that trust is affected by AI performance, we designed high-performance AI (HPAI) systems and low-performance AI (LPAI) systems. Each participant interacted with an HPAI and an LPAI system for 20 scenarios each. Participants will be randomly assigned to interact first with either the HPAI or LPAI systems. There will be a break

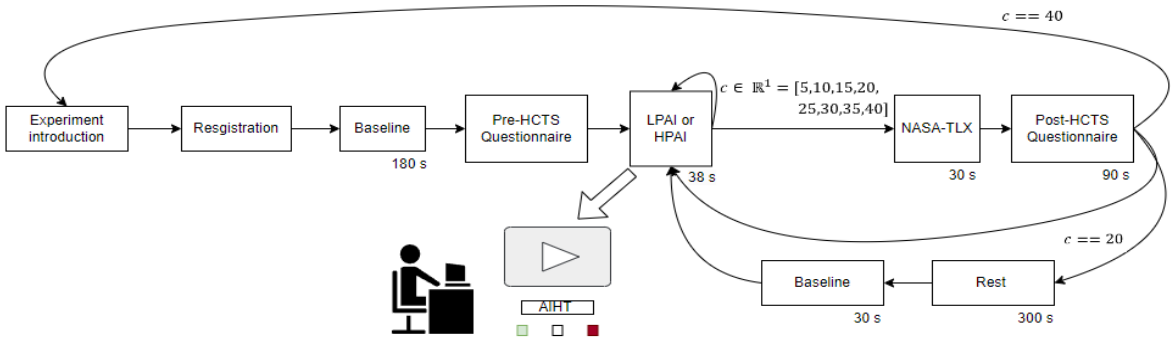


Fig. 1. Experience sequence for cognitive performance and trust assessment on the AIHT Platform.

between the interaction with the first AI system and the second AI system. As each scenario lasts approximately 38 seconds, the total procedure takes about 60 minutes, accounting for training and downtime between scenarios.

A. AIHT Platform Implementation

The AIHT platform design follows a simplistic two-layer web architecture topology [25]. The web architecture topology combines three components of application, presentation, processing, and database in two modalities on a machine with specifications Intel(R) Core (TM) Ultra, 32 Gb RAM, 3.8 GHz processor speed, Windows 11 operating system, Apache 2.4.54 server and Oracle Database 18c Express Edition Production. Bootstrap 5 framework containing user interface (UI) components such as HTML, CSS, JavaScript, AJAX and PHP were used to render and control logic of the experiments presented to the participants and investigators. Below is the pseudocode for the AIHT platform.

```
START Experiment
SET numberOfExperiments = 40
SET count = 0
WHILE count < numberOfExperiments DO
   Play vignettes
   AIHT offers recommendations
   Participant decides and gets a response
   IF count is in [5, 10, 15, 20, 25, 30, 35, 40]
   THEN
      IF count == 20 THEN
            Participant fills NAS-TLX
            Participant fills HCTS
            Participant takes a break
      ELSE IF count == 40 THEN
            Participant fills NAS-TLX
            Participant fills HCTS
            Start new experiment
      ELSE
            Participant fills NAS-TLX
            Participant fills HCTS
      ENDIF
  ENDIF
      INCREMENT count
   ENDWHILE
```

B. Physiological Recording

The Empatica Embrace Plus watch is used for acquisition of EDA. It has a range of 0.01 – 100 microsiemens, resolution of 1 digit - 900 pico Siemens and at a sampling frequency of 4 Hz.

C. AIHT Platform Validation

In our ongoing study, we aim to recruit nursing students, at sophomore, or junior level. Individuals who use stimulants such as caffeine will be excluded. For preliminary validation of this platform, we are presenting the data of one participant to show the efficacy of the AIHT platform. Upon arrival at the lab, the participants were informed about the study's purpose and given consent forms. After providing consent, they registered on the AIHT platform. This study was conducted under an approved IRB protocol (B2024-0023) at the University of Connecticut.

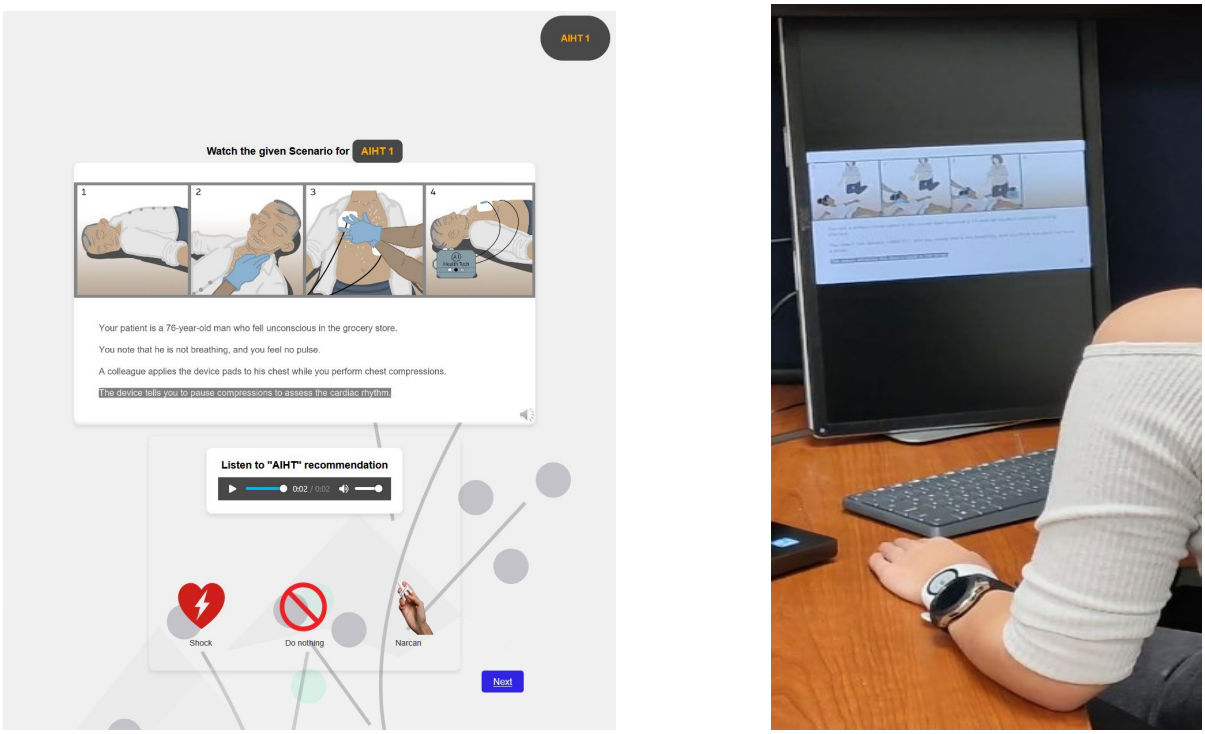
III. RESULTS

A. Signal Processing

This study analyzes data comprised of acquired EDA signals, trust-influenced reaction times, and cognitive performance of trainee nurses during experiments Fig. 2. Raw EDA signals were processed using a 5-second window median filter for smoothing, followed by a low pass FIR filter of 1 Hz to remove high-frequency components. The cleaned signal was then decomposed into tonic (slowly varying) and phasic (rapidly changing) components using the cvxEDA technique [26]. The EDA response from a participant during interaction with a HPAI is shown on Fig. 3. The HPAI has 95% of accuracy in their recommendations. During this phase, an average of 34 EDA peaks were observed, and the participant exhibited a 90% of accuracy in their responses. The average reaction time was 4.75 seconds. This suggests that the AIHT system's high accuracy fostered trust, leading to strong performance (Fig. 3a).

After a five-minute rest period, the participants interacted with a LPAI with an accuracy of 60% in their recommendations. This phase resulted in a 30% increase in

This study was funded by the *NursEng Healthcare Innovation Seed Grant*, Nursing and Engineering Innovation Center, University of Connecticut, and the *Faculty Pilot Grant Program*, Penn Nursing, University of Pennsylvania.



(a) (b)

Fig. 1. Experimental view of the AIHT platform during experiments (a) Instance of a scenario illustrated with vignettes b) Trainee Nurse using the AIHT platform.

EDA peaks compared to the trust-building phase, potentially indicating increased cognitive effort due to the AIHT's reduced reliability. Simultaneously, performance accuracy of the participant dropped by 50%, suggesting that diminished AI accuracy weakened trainee nurses' trust, negatively affecting their performance (Fig. 3b).

IV. DISCUSSION AND CONCLUSION

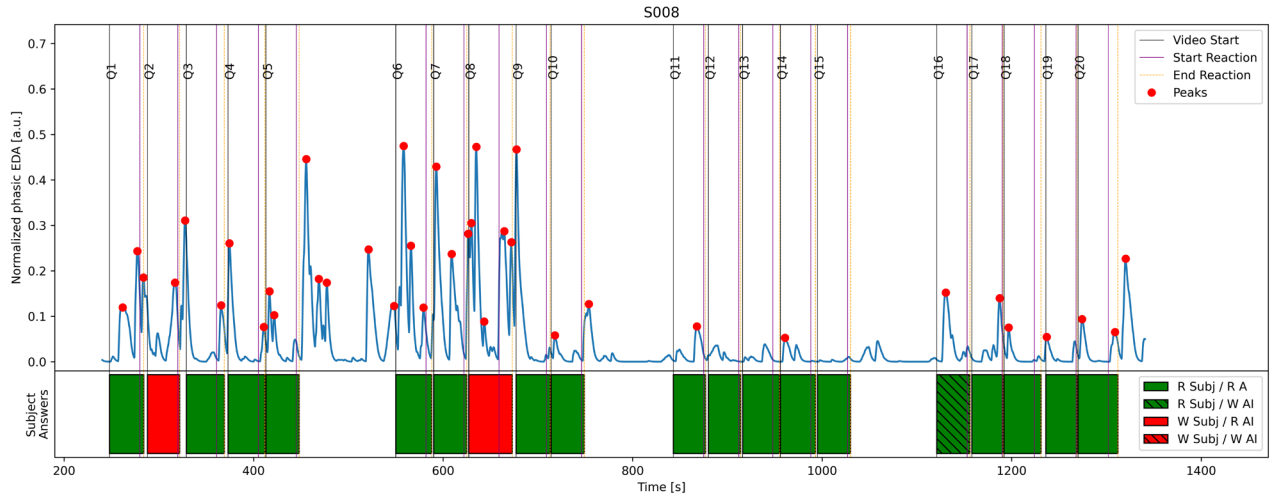
This preliminary study introduced a novel platform for assessing trainee nurses' trust in AI healthcare technology systems. While our findings are based on limited initial data, they suggest that the integration of physiological signals such as EDA with behavioral measures may offer valuable insights into trust dynamics in healthcare AI interactions. The observed differences in EDA response patterns, reaction times, and accuracy between interactions with high-performance and low-performance AI systems align with current understanding of trust formation [11], [12], [13]. These preliminary observations align with previous studies exploring physiological markers of trust and warrant further investigation with a larger sample to establish their reliability and generalizability [18], [19], [20], [21].

Our work has several limitations that should be acknowledged. Most significantly, the current validation is based on data from a single participant, which serves as proof of concept rather than definitive evidence. The planned expansion to 40 participants will provide a more robust evaluation of the platform's efficacy. Additionally, the artificial laboratory setting may not fully capture the complexity of real-world clinical decision-making environments that trainee nurses will encounter [1], [7]. Despite these limitations, this study represents an important step toward developing objective measures of trust in healthcare AI. The AIHT platform's integration of realistic video scenarios with physiological monitoring offers a promising approach for future research and potential applications in nursing education [8]. As AI systems become increasingly prevalent in healthcare settings, understanding the factors that influence appropriate trust calibration will be essential for effective human-AI collaboration [3].

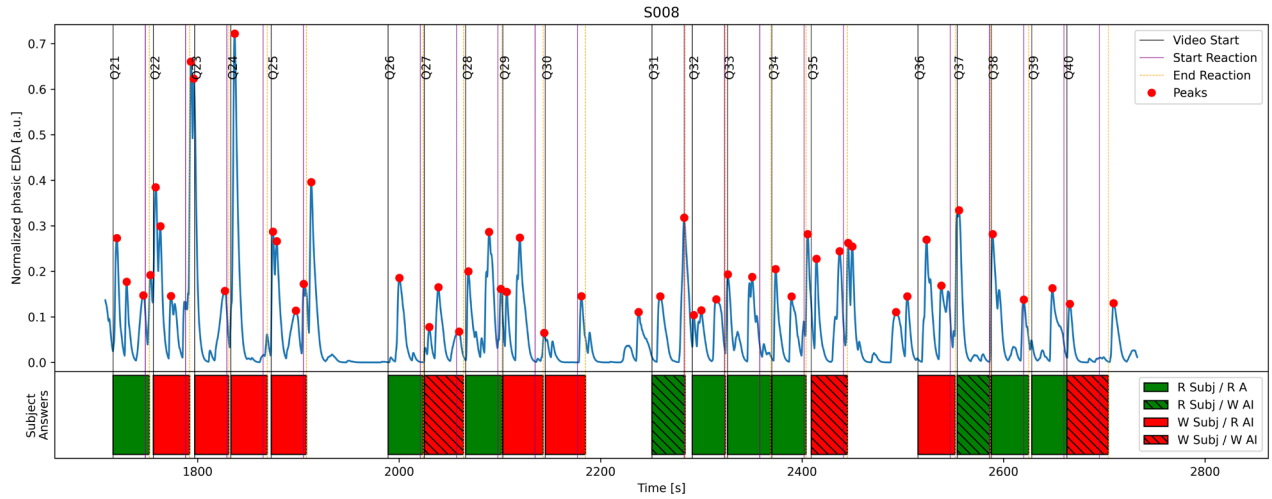
Future work will focus on validating the platform with a larger sample and exploring personalized difficulty adjustments to optimize learning outcomes. We also aim to investigate how different AI explanation styles might influence trust formation and decision quality in healthcare contexts, building on existing frameworks for AI integration in health professions education [1], [3].

ACKNOWLEDGMENT

Authors thank Teresa Graziano, Valorie MacKenna, and Katie Tomasko for their support in the design of the scenarios.



(a)



(b)

Fig. 2. Plot of phasic components, reaction times and performance accuracy for an AIHT experiment. a) high-performance AI (HPAI) interaction b) low-performance AI (LPAI) interaction

REFERENCES

[1] B. C. Gin *et al.*, "Entrustment and EPAs for Artificial Intelligence (AI): A Framework to Safeguard the Use of AI in Health Professions Education," *Acad Med*, Nov. 2024, doi: 10.1097/ACM.0000000000005930.

[2] S. Rasouli, D. Alkurdi, and B. Jia, "The Role of Artificial Intelligence in Modern Medical Education and Practice: A Systematic Literature Review," Jul. 26, 2024, *medRxiv*. doi: 10.1101/2024.07.25.24311022.

[3] S. Labkoff *et al.*, "Toward a responsible future: recommendations for AI-enabled clinical decision support," *Journal of the American Medical Informatics Association*, vol. 31, no. 11, pp. 2730–2739, Nov. 2024, doi: 10.1093/jamia/ocae209.

[4] K. Gupta, R. Hajika, Y. S. Pai, A. Duenser, M. Lochner, and M. Billinghurst, "In AI We Trust: Investigating the Relationship between Biosignals, Trust and Cognitive Load in VR," in *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, in VRST '19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1–10. doi: 10.1145/3359996.3364276.

[5] W.-L. Hu, K. Akash, N. Jain, and T. Reid, "Real-Time Sensing of Trust in Human-Machine Interactions**This material is based upon work supported by the National Science Foundation under Award No. 1548616. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.," *IFAC-PapersOnLine*, vol. 49, no. 32, pp. 48–53, 2016, doi: 10.1016/j.ifacol.2016.12.188.

[6] G. Campagna, D. Chrysostomou, and M. Rehm, "Investigating Electrodermal Activity for Trust Assessment in Industrial Human-Robot Collaboration: 21st International Conference on Ubiquitous Robots (UR 2024)," *2024 21st International Conference on Ubiquitous Robots, UR 2024*, pp. 880–885, Jul. 2024, doi: 10.1109/UR61395.2024.10597471.

[7] Y. S. J. Aquino *et al.*, "Utopia versus dystopia: Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills," *International Journal of Medical Informatics*, vol. 169, p. 104903, Jan. 2023, doi: 10.1016/j.ijmedinf.2022.104903.

[8] M. Masoumian Hosseini, T. Masoumain Hosseini, and K. Qayumi, "Integration of Artificial Intelligence in Medical Education: Opportunities, Challenges, and Ethical Considerations," *J Med Edu*, vol. 22, no. 1, Jan. 2024, doi: 10.5812/jme-140890.

[9] F. Li and S. Betts, "(PDF) Trust: What It Is And What It Is Not," *ResearchGate*, Oct. 2024, doi: 10.19030/iber.v2i7.3825.

[10] B. Alhaji, S. Büttner, S. Sanjay Kumar, and M. Prilla, "Trust dynamics in human interaction with an industrial robot," *Behaviour & Information Technology*, vol. 44, no. 2, pp. 266–288, Jan. 2025, doi: 10.1080/0144929X.2024.2316284.

[11] K. A. Hoff and M. Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust," *Hum Factors*, vol. 57, no. 3, pp. 407–434, May 2015, doi: 10.1177/0018720814547570.

[12] C. A. Hill and E. A. O'Hara O'Connor, "A Cognitive Theory of Trust," Dec. 01, 2005, *Social Science Research Network, Rochester, NY*: 869423. doi: 10.2139/ssrn.869423.

[13] K. Kostick-Quenet, B. H. Lang, J. Smith, M. Hurley, and J. Blumenthal-Barby, "Trust criteria for artificial intelligence in health: normative and epistemic considerations," *Journal of Medical Ethics*, vol. 50, no. 8, pp. 544–551, Aug. 2024, doi: 10.1136/jme-2023-109338.

[14] A. Jaiswal, M. Z. Zadeh, A. Hebri, and F. Makedon, "Assessing Fatigue with Multimodal Wearable Sensors and Machine Learning," Oct. 25, 2022, *arXiv*: arXiv:2205.00287. doi: 10.48550/arXiv.2205.00287.

[15] Y. Kong, H. F. Posada-Quintero, and K. H. Chon, "Real-Time High-Level Acute Pain Detection Using a Smartphone and a Wrist-Worn Electrodermal Activity Sensor," *Sensors*, vol. 21, no. 12, Art. no. 12, Jan. 2021, doi: 10.3390/s21123956.

[16] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 2, pp. 410–417, Mar. 2010, doi: 10.1109/TITB.2009.2036164.

[17] A. Khawaji, J. Zhou, F. Chen, and N. Marcus, "Using Galvanic Skin Response (GSR) to Measure Trust and Cognitive Load in the Text-Chat Environment," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, in CHI EA '15. New York, NY, USA: Association for Computing Machinery, Apr. 2015, pp. 1989–1994. doi: 10.1145/2702613.2732766.

[18] I. B. Ajenaghughrure, S. D. C. Sousa, and D. Lamas, "Measuring Trust with Psychophysiological Signals: A Systematic Mapping Study of Approaches Used," *Multimodal Technologies and Interaction*, vol. 4, no. 3, Art. no. 3, Sep. 2020, doi: 10.3390/mti4030063.

[19] K. Akash, T. Reid, and N. Jain, "Adaptive Probabilistic Classification of Dynamic Processes: A Case Study on Human Trust in Automation," in *2018 Annual American Control Conference (ACC)*, Milwaukee, WI: IEEE, Jun. 2018, pp. 246–251. doi: 10.23919/ACC.2018.8431132.

[20] L. Cominelli *et al.*, "Promises and trust in human–robot interaction," *Sci Rep*, vol. 11, no. 1, p. 9687, May 2021, doi: 10.1038/s41598-021-88622-9.

[21] F. Walker, J. Wang, M. H. Martens, and W. B. Verwey, "Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 401–412, Jul. 2019, doi: 10.1016/j.trf.2019.05.021.

[22] J. Zhou, H. Hu, Z. Li, K. Yu, and F. Chen, "Physiological Indicators for User Trust in Machine Learning with Influence Enhanced Fact-Checking," in *Machine Learning and Knowledge Extraction: Third IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2019, Canterbury, UK, August 26–29, 2019, Proceedings*, Berlin, Heidelberg: Springer-Verlag, Aug. 2019, pp. 94–113. doi: 10.1007/978-3-030-29726-8_7.

[23] S. Gulati, S. Sousa, and D. Lamas, "Design, development and evaluation of a human-computer trust scale," *Behaviour & Information Technology*, vol. 38, no. 10, pp. 1004–1015, Oct. 2019, doi: 10.1080/0144929X.2019.1656779.

[24] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006, doi: 10.1177/154193120605000909.

[25] B. E. Faremi, J. Stavres, N. Oliveira, Z. Zhou, and A. H. Sung, "Enhancing Machine Learning Performance with Continuous In-Session Ground Truth Scores: Pilot Study on Objective Skeletal Muscle Pain Intensity Prediction," unpulished Aug. 02, 2023, *arXiv*: arXiv:2308.00886. doi: 10.48550/arXiv.2308.00886.

[26] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing," *IEEE Trans Biomed Eng*, vol. 63, no. 4, pp. 797–804, Apr. 2016, doi: 10.1109/TBME.2015.2474131.